

**Webinar Transcript:**  
**American Community Survey: Administrative Record Research**  
**May 13, 2020**

Coordinator: Welcome and thank you for standing by. All participants are now in a listen-only mode until the question-and-answer session of today's presentation. This call is being recorded. If you have any objections, you may disconnect at this time, and I would now like to turn the call over today's host, (Jennifer Ortman). Thank you, you may begin.

(Jennifer Ortman): Thank you, (Charles). Good afternoon, everyone, and thank you so much for joining us. I think we may have record-breaking attendance, so wherever you're joining us from, I hope you're all staying safe and healthy, and hopefully this will be a nice break from whatever you were doing this afternoon. So I'm (Jennifer Ortman), and it's my pleasure to kick off today's webinar on research to integrate administrative data into the American Community Survey. We have an exciting lineup of presentations, but first, I'd like to set the stage.

So the ACS is the nation's most current, reliable, and accessible data source for local statistics on critical planning topics. We survey 3.5 million addresses and inform over \$675 billion of federal funding each year. We provide data on more than 40 topics, and relay 11 billion estimates each year. ACS data are a crucial resource when our nation's confronted with national disasters, pandemics, and other major events. The ACS provides data on a variety of topics. We ask about several characteristics of the population, covering social topics, such as educational attainment and language use, demographic characteristics, such as age and sex, and economic characteristics such as

employment and income. We also ask a number of questions about the characteristics of the nation's housing units, covering things like computer and internet use, utilities, and home value.

The ACS is particularly useful because it provides data for very detailed geographic levels, including the nation, states, counties, census tracts, and block groups. This is the only federal survey to provide data for small geographies, making it uniquely suited for use by policymakers, planners, and others who seek information at the community level. We first attempt to collect ACS data through self-response, asking respondents to either go online to complete the survey or complete it on paper, which they mail back to us. We then conduct follow-ups using personal interviews to solicit responses from those who did not self-respond. Our program continually evolves to ensure the ACS program provides quality data to our many data users, while also providing a quality experience for our respondents. As we look to the future, the next phase of evolution means embracing the potential and power of administrative data.

There are several reasons why we think administrative data may enhance the ACS program. This includes increasing data quality by filling in missing responses and using administrative data to evaluate and enrich survey data. We can save time and improve the respondent experience by reducing the number of questions asked on the ACS. We can provide cost savings by identifying vacant housing units and reducing the need for follow-up visits. And, last but not least, the Census Bureau is mandated by Title XIII of the US Code to use already available information.

Over the last several years, we have made great strides to determine the feasibility of integrating administrative data in various aspects of the ACS life cycle. You will hear some highlights of the various stages of this research,

which will give you a foundation for the research projects you will hear more about today.

First, we have to identify the available data sources. The Census Bureau has access to a variety of administrative data sources, including federal data, which come from agencies such as the IRS and Social Security Administration, state and local data, including data about programs, such as TANF and SNAP, third party data, which comes from third-party vendors who compile and sell data. This includes data from Black Knight and DSGI. Next, we evaluate each data source using a set of guiding principles to determine (inaudible).

The principles cover things like conceptual alignment, which is the extent to which the administrative data corresponds to the concept the ACS intends to measure, and quality, which is whether the administrative data meets the quality requirements for our published data. Using these principles, we identify the best candidates for a given research endeavor, whether it is replacement, allocation, or some other use. Here, you see results from our evaluation of data about housing units, which indicated the year built, acreage, property value, and property tax were the most promising for the task of replacing survey data with administrative data.

We then move forward with research to experiment with the specified use of the administrative data to either identify next steps for research or confirm feasibility to recommend implementation in ACS progression systems. Examples of current research endeavors include predicting vacant housing units to reduce need for follow-up visits, replace survey responses with administrative data to reduce the number of questions asked of the respondents, and filling missing values to improve data quality, enriching ACS data linkages to administrative data to expand the number of topics

covered, and using administrative records to improve measurement of income.

So, hopefully that gives you all a bit of background to understand why the ACS program is interested in using administrative data, and the ways in which we think we might be able to leverage this valuable resource. We have many exciting projects underway, and now it's time for you to hear more from the researchers who are carrying out this work. First, you're going to hear from (Andy Keller) about work he's doing to predict vacant housing units in the ACS. Next, (Chase Sawyer) is going to talk about his work to analyze differences between administrative data and survey responses about property value. Third, (Sandy Clark) will talk about her work to use alternative data to fill in missing guidance for demographic characteristics, and (Leah Clark) will wrap things up by discussing her work to make ACS and IRS data to assess college attendance and completion.

(Andy Keller): So, this is (Andy Keller). Thanks for your time this afternoon, and I'm going to talk about some initial research predicting vacant units in the ACS, as (Jennifer) said. So, moving on to the next slide, here's a brief outline. First, I'm going to talk about related work predicting vacant units and how that has been implemented. Next, I will discuss how the ACS-related research has progressed thus far, including the data used and the methods used to -predict the vacant units.

I will then spend the majority of the time going over a simulation on previous ACS data, first generating a model and then evaluating the quality of the predictions that are generated from that model. And then finally, I will finish up with what we have learned thus far and possible alterations or improvements to the research as we go forward.

So for the 2020 Census, a major research (emphasis was to) use administrative

records to inform the non-response follow up operations. So (along that vain), we developed models to identify occupied, vacant, and delete addresses, and delete addresses are those that did not reach the Census Bureau's definition of a housing unit. That information has been used to modify the contact strategy for addresses for which we have high confidence in any of the three statuses. For example, we can conduct one visit at units with high confidence of final status, rather than continued visits. The idea of this research on ACS was to apply a similar concept to predict units with high confidence of vacancy, and I should mention there's conceptual difference in that the census measures point in time vacancy.

But the overarching idea here consists in that we use the vacancy probability to inform the contact strategy or the sampling strategy in the case of ACS. So we developed a model where the dependent variable is vacancy outcome. We used the previous year, or we can use years of ACS data to form the training data for our model. So that data includes administrative record data of the same vintage or similar vintage, operational data that we gained from the survey, address-level information from our master address (inaudible) maintained by the Census Bureau, and information at the block group level, taken from the ACS planning database.

So the main idea of the methodology is to fit our model over the previous year or years and then apply the parameter estimates to our current vintage of ACS data as external validation data. We then have a predicted probability of vacancy for every ACS unit in the mailable, computer-assisted personal interview universe. So, what data are we using in our model? Let's go a little bit more into that. So, with respect to administrative records, we use aggregated public information purchased by the Census Bureau consisting of local tax, deed, and mortgage information. We also use information concerning land use or absence of an owner at the address or ownership rights

on the unit. We use third-party AR data, providing information about persons at that address, as well as national change of address information from the United States Postal Service. And you can note here that some of the variables are more helpful prediction. With respect to ACS operational data, we use mailing operations.

So this is our undeliverable as addressed data from the United States Postal Service, as well as indication of vacancy from internet response. And as you might expect, these two variables are particularly powerful information because of their timeliness. We also get address-level information from the Master Address File. This includes information from the Delivery Sequence File, which indicates whether the address is residential, commercial, or excluded from the delivery (inaudible) statistics computed by the United States Postal Service. We also use information covering the type of housing unit, whether it's a single unit, a multi-unit, a trailer, or other type of unit. We also use information covering the type of delivery point. Last, we have block-level group indicators, such as the poverty rate, the rental rate, other language rate and Hispanic rate for area-level information that informs our model.

So, let me give into some specifics with respect to this simulation I'm going to talk about. So we take the mailable computer-assisted personal interview (CAPI) universe cases from the 2016 and 2017 ACS universe. We fit the model on the 2016 data, using the vacancy outcome as our dependent variable. We apply the perimeter estimates to 2017 ACS universe, so that we have predicted vacant probabilities over the entire universe of cases, and then we sort the probabilities from greatest to least so at the top are the cases that we believe are the most likely to be vacant via our model.

So we iterate over the top percentages by picking a threshold - say the top ten percent or top five percent of predicted vacant probabilities. We then compare

the actual results to the actual results in 2017, to see how many were indeed vacant, and note that about 25 percent of the mailable CAPI cases were vacant in the 2017 ACS. So, taking a step back, there are really two research questions inherent in this project. The first was, well, can we actually build a model that can reasonably identify the vacant cases? And then the second gets into –well- what percentage of our threshold of those should be used to form a treatment? So I have a histogram of these predicted probabilities right here that I think somewhat addresses both of those questions.

So for the first question the separation of the humps - where we have this kind of big hump on the left, and you have almost two smaller humps on the right - his a positive sign that a model can be built in the sense that the vacant cases can be separated out from the non-vacant cases. So the non-vacant cases are on the left, and the vacant cases are on the right.

And for the second question, as we move from right to left, maybe it's time to stop after the first hump, moving right to left, or after the second hump, taking into account the cost-benefit trade off. So you kind of have to figure out, in terms of the threshold, where should I stop in terms of - using the cases for some sort of treatment? So here's an example where I take the top ten percent of vacant predicted probabilities. I then look at the actual results from the 2017 ACS. So for each state, the denominator is (inaudible) the number of cases within that state that fell within the top ten percent, and the numerator is the number of those cases that were vacant. Then we can calculate a true positive rate for each state.

So for this particular graph, the darker colors represent those states with the higher true positive rates. You can see Wisconsin, Michigan, Florida, New Jersey. And some of the lower rates are are the ones with the lighter colors. For example, Oklahoma, West Virginia. So I thought this was an interesting

plot in the sense that the false positive rate is not uniform across the states. I also think this bears some further investigation on my end, as to understanding the lack of uniformity. For example, is it due to the variance in the coverage of administrative data? So, this next slide, with the bar graph shows the distribution of outcomes for the various thresholds as we move from the left, over the top .5 percent, to the top ten percent on x-axis.

So, each bar is a stacked plot, where the percentages of each bar add to 100 percent. The light green at the bottom of each vertical bar is the percentage of cases where the outcome is vacant. So again, this can be thought of as the true positive rate. The light blue at the top of each vertical bar is the percentage of cases where the outcome is occupied, and this is the result that is the most concerning, because we would be identify cases that are vacant, that are actually indeed occupied, and risk missing interviews from those people. The medium blue in the middle of each vertical bar is the percentage of cases where the outcome was not a housing unit, so we don't necessarily miss people, but we do miss the fact that it was a housing unit.

So a couple of things to notice. As you move from left to right on the bars for the previous slide, for the threshold of vacant cases, more of the results are occupied – so you can see less green as you move from left to right, and this also can inform the tolerance for picking a threshold. So moving forward, on an earlier slide, we saw there's some variation across states, of the true positive rate given that (inaudible) ten percent threshold. So each state has a certain distribution of the mailable CAPI universe. And depending on the cost benefit, we can pick a threshold, as we said earlier, of predicted vacant probabilities to treat. And what I've seen is, regardless of the top threshold we select, the state-level distribution of the top threshold is not the same as the overall mailable CAPI universe's state distribution.



For example, say state X is three percent of the mailable CAPI distribution. You pick your top ten percent of predicted cases, and the state may be seven percent of that top ten percent. The idea here is that there's a differential aspect of picking out a threshold, and the implications need to be thought through.

So another thing I wanted to look at was the implication on the occupied cases that were identified as vacant to see the effect on traditionally undercounted groups and the three groups I looked at were Hispanic, Black, and 0 to 4. So among occupied mailable CAPI cases, about 25 percent have someone who's Hispanic, about 21 percent have someone who's Black, and about 15 percent have someone who's aged 0 to 4 (inaudible). And so for each threshold, I looked at the occupied units that we identified to see the composition of those units. And in this graph, Hispanic is blue, Black is orange, and the 0 to 4 group is red. So, across all thresholds, the percent of occupied units with someone who's Hispanic is around ten to 12 percent, and the percent of occupied units with someone who's Black is around 15 to 17 percent, and then with the red - the percent of occupied units with someone who is 0 to 4 is around ten to 12 percent.

And so what we're learning from this is that the occupied units that we've identified as vacant are proportionately fewer of these undercounted groups within them. So moving forward, another kind of exercise I've been doing, as I've been going through this modeling process, is I've been interested in studying where we have high vacant predicted probabilities where the outcome was non-vacant.

And the goal is to get a sense of the cases where we might be concerned about calling them vacant. And this is somewhat an iterative process because I turn around and then use this information to create a better model. So, it's kind of a

give and take, as I work through the model. So I use the decision tree to create business rules for a combination variables for which a non-vacant outcome occurs most frequently, given a certain threshold. So for example, I take the top ten percent of predicted cases. An example case where I'm misclassifying is addresses where there's no land use indicated on administrative records, and the address is not on the (inaudible) Delivery Sequence File of the previous fall. So in this case, it was about 15 percent of the universe and 52 percent of the cases were not vacant. So again, I'm using this this kind of analysis of false positives to inform my model and refine as I go forward, which is almost like a give and take.

So moving forward, with my conclusion- so the goal of this processes we work through it, it's to see if we can model vacants in the ACS universe using a combination of address-level, ACS operational, geographic, and administrative records information. And so the cost-benefit analysis is used to help determine the threshold for using the best predictions. Generally, cases within the top thresholds contain relatively fewer numbers of the harder to count groups. We do observe a difference of true positive rates across the states. As we're doing this, we developed business rules to identify false positive cases which can help inform and refine the models, and I should note that not all false positive cases are occupied units, some are actually addresses that are not housing units..

So, (inaudible) just kind of a further discussion on ideas. This is initial (inaudible) research in the sense that different aspects of this research can also be tweaked. For instance, this research universe doesn't have to be restricted to mailable CAPI cases. It can include all CAPI cases though doing so would probably necessitate rethinking the model since the mailable CAPI model uses information, UAA information (undeliverable as addressed information) as a good predictor. So, if we are going to have different universes, I need to

rethink and readjust the models, or develop another model. Another aspect is making this model more adaptive. So, we have these initial vacant probabilities of our CAPI universe. We can send all those cases out to the field and update the probabilities with that information from the visit, so we can use our visits to update and refine our model. That's kind of a dynamic adaptive idea.

So we can use the predictive probability to alter the contact strategy, and this is in fact how we're using this type of idea in the 2020 Census. And finally, we can also use the predicted probabilities to alter the sampling rate, particularly in ACS-so we can change from a higher sampling rate to a lower sampling rate. The risk here is that occupied cases would inflate variances. But the good kind of thing from this research is a lot of these things can be simulated with different parameters at your desk so you can learn a lot just by playing with the different parameters. So, thank you for your time.

(Chase Sawyer): Thank you, (Andy), for talking over some of the ways that the ACS may be able to use administrative records to utilize in their field operations. I'm (Chase Sawyer), and I'm excited to talk with everyone today about some more work that we're doing, regarding administrative records. In this instance, we're actually looking at how the use of property tax records could be utilized in the American Community Survey and how those may actually affect our estimates, and we're going to specifically be looking at property tax data.

So, to set the stage here, a few years ago, we did a research project where we used housing administrative record data to simulate how that would affect ACS estimates, if we were to do so.

So, our goal in this project was to see how we could use administrative records, what the impact on our estimates would be, and then just also learn

what we would need to do in the production system to make this happen.

To do this, we used the 2015 ACS responses, and today, I'm going to focus on one of these four promising housing items that we had, which was property value. And to do this, we used a modeled administrative record data from Core Logics' automated valuation model, to replace respondent-provided values. And we ended up creating a simulated version of the ACS estimates to compare to the published 2015 ACS estimates.

And so, as part of this, we went ahead and we created an adaptive design. We didn't use every single administrative record that we had available for us, or were able to link, because we wanted to see how this would work in production.

And so, to do this, what we did was we used the ACS sample, and we assumed that people that responded from an internet, a computer-assisted telephone interview, or using - or as part of the personal interview follow-up, that these instruments we would be able to adjust, so that if we knew we had administrative record available, that we could go ahead and skip that question in the instrument. If the respondent, though, provided a mailed survey response, we assume that we would need to use that response out of respect to the fact that they took the time to provide us a response, and we did this because we assumed that we wouldn't be able to have it be feasible to have multiple versions of the mailed questionnaire.

So, we would just ask the question - all of the questions that we had of the mailed respondents. If the response came back to us, though, and it was blank, instead of taking it and pushing it through to our imputation process, we went ahead and used the administrative record, if it was available. And so now, as I get into some of our research results for property value, I do just want to

remind you that these are simulated estimates for the ACS, and so it's not a direct comparison between the respondent-provided value and the administrative record value.

And also, as we start to dive into some of these different crosses that we did to the variables, I want to point out that the results may be confounded by whether or not we were able to link administrative records to that housing unit. So, an example of this is, like, it may be easier for us to link our administrative record data to certain households than it is others. And so, if that's a characteristic that makes linkage difficult, then it may also be where we see some of these simulated estimates, that the differences aren't as large.

Our key measures for property value in this simulation were a median property value, the number of properties that were reported as being worth less than \$10,000, and the number of properties that were reported to us as being worth \$2 million or more, and we chose these properties because we wanted to get an overall sense of the median property value, but also look at some of these. These are values that are on the edge of the bins that we published, so we wanted to see if maybe we're getting a smaller number of these different outliers. The first results, though, that I want to review with you today involve our medium property value. And so, overall, you see that there was a difference between our simulated estimates and our published estimates of 6.3 percent. When we look at this geographically, though, there are differences throughout the different states. So we can see that most of the states fell on either side of this bin, and they're pretty evenly split, with about five states having differences of nine percent to 19.9 percent, and some that were closer to not being statistically different. But all of the states saw a decrease in the median property value between our simulated and our published estimates.

What you'll see here, though, on this next slide is that this is fairly related to the burden reduction or the number of administrative records that we're actually able to swap out from our respondents' answers. So, what we're seeing here is that the respondent values are usually lower than the records that - or I'm sorry, the respondent values are actually higher than the records that we're getting from the administrative records. And that isn't distributed equally throughout the states, because our ability to link to administrative records data isn't distributed equally throughout the states. And so, this is a big deal for us with the American Community Surveys, because our estimates are used to fund a great deal of block grants and public funding. So we need to make sure that the results of our using administrative records are done equally. So I also went ahead and broke out some additional data points for us to review today.

So, I have a few more median property value for us to look at. We're going to look at how the differences, based on when the property was built, when the person moved in, as well as the mortgage status, and then we're also going to spend a little bit of time looking at the distribution of property value and, in the case of households that are worth less than \$10,000, we'll look at the household income in those homes. So here, we can see that when we look at the differences of median property value by the year built, there's not like a discernible linear pattern here between the year that the unit was built and when - or - and the difference with the property value. What we do see, though, is that the largest difference for median property value was with homes that were built before 1940.

And so, this is something that we need to look into a little bit further, to see why these housing units had such a large difference between those estimates. Another breakout that we have is for the year that the person moved into the household for their property value, as you can see, there is a kind of sloping

decrease here, based on when the person moved into the unit and when - or how much there was difference with that median property value. People that moved into the home before 2000 had a greater decrease, and people that moved in after 2000 had a smaller decrease.

And some reasons we kind of think that this may be is that when we're looking at the data, we see that people that are younger are more likely to over-value - well, are more likely to have a higher value on their home than those that have been in there longer. And as we're actually going to look at in the next slide, they may have reasons regarding their mortgage. So, this last breakout that I have for median property value looks at the difference by mortgage status, and we see that units that have a mortgage were likely to have a decrease that was higher than the national average, whereas units without a mortgage had a lower decrease in the amount of administrative records.

And so, a reason why this may be is, as I thought this through, is that people that have a mortgage, they want their homes to be kind of worth as much as possible, maybe more than someone that doesn't have a mortgage, because that's showing the difference between what they owe on the home and what the home is worth. And this also is kind of a confounding influence with what we're talking about with age and when people move into their homes. So, this is definitely something that we'd like to research a little bit more and look into some of these and see why the median property value may be different for these characteristics.

So to give an overview of the breakouts for median property value, we saw that for year built, there wasn't a discernible pattern, as we went through our different categories. But units built before 1940 had the largest decrease. If we looked at year moved in, units that have been moved into more recently have

the largest decrease as opposed to those that someone has lived in for a while. And also, we're seeing that when we look at mortgage status, the decrease is largest for units that currently have a mortgage.

So there was these other two measures that we thought were important for us to look at, as we did the housing simulation, and this was properties that were valued less than \$10,000, and properties that were valued at 2 thousand- sorry, \$2 million or more. And we saw decreases for those categories of 16.3 percent and 26.6 percent, respectively. On this next slide, though, what I've done is I've taken all of the bins that we have for property value to try to look at them and see what happens with that distribution and what the differences were across the sample. And so, what we're seeing here in our simulated estimates are that kind of our outliers are where we had changes, whereas homes that were worth between \$50,000 to about \$250,000 are more - are where we're seeing the gains. So, I want to dig into this a little bit further, so we can actually confirm this, but to look and see if those homes that were worth less than \$50,000 and so they're kind of feeding into the lower end of where we're seeing the gains, and then the two - more than \$300,000 or more, and see if those are kind of the upper end of the gain.

But actually, to look into this and see where people are moving on this continuum. Another table that we have, that we're able to go ahead and look at is the property value of less than \$10,000. And so, we actually have a breakout in the American Community Survey that is we have a table that shows the household income for these properties. And so, on this next slide, you'll see where I've gone and taken these homes and I want to see where kind of they lined up, and what areas - or what groups, based on their household income, had the largest decreases.

And so you can see here, on the lower end of household income, there's a



smaller decrease, but there is a decrease in the number of these homes. And then, if you look on the higher end, you can see that the increase just gets larger and larger. So, the one result here that stands out to me is that we would - if we had used this methodology in the ACS in 2015, half of the homes that were valued at less than \$10,000, that people told us their household income was worth more than \$100,000, we would decrease that by 50 percent almost.

And so, this is something for us to look into a little bit more, as we're looking at cleaning data and doing edits and checking on data to see if some of these values make sense, because it is - while we are trying to measure small areas in ACS, it's probably fairly unlikely that there are a lot of people that earned more than \$100,000 a year, but have a home that's worth less than \$10,000. So some of the next steps that we're hoping to do with this research is look at comparing ACS responses with a new set of administrative record data.

We also want to do a little bit more research to look specifically at some of these housing characteristics and not just look at our estimates, but look at how these are different from each other, based on the actual responses that we received, versus the administrative records. And we're also looking to do modeling projects on our own, to start to look at could we use modeling of some from this information to determine when we need to look at making edits, like in the instance of the high household income but low property value, as well as looking at creating our own automated valuation models for homes so that we would be able to have the methodology there and understand it.

So these are two of the research products that we have on this work. We'll have the hyperlinks again later on in the slide, so you can see them, but there's a report that's been published about this information, as well as we have a data visualization that lets you look not just at the property value information that

I've gone over today, but also look at some of the other topics that we looked at with this work. So I think with that, I'll go ahead and turn the time over to (Sandy Clark), so she can talk to us about some of the work we're doing with demographic data.

(Sandy Clark): Thanks, (Chase). Those are some very interesting findings, and also I guess we have some more, to put into some other research projects that you can look into and hopefully will be able to do more webinars and provide some of these results in the future. So, good afternoon, everyone. My name is (Sandy Clark). Today, I will discuss research using alternative data sources to fill in missing values for demographic characteristics in the ACS.

As (Jennifer) mentioned earlier, there are several reasons for using administrative data in surveys such as the ACS. (Andy) discussed using these data to predict the vacant housing units before going into the field to reduce the need for follow-up visits, and therefore save money. (Chase) studied using administrative data to replace the need to ask survey items, and also possibly model results in the future. This would reduce respondent burden. My research tests using administrative data to impute data for survey items left blank by respondents. This improves data quality.

All of the research discussed today puts already available data to good use, satisfying Title XIII of the US Code. There's a good deal of research to support using administrative data in lieu of statistical procedures such as the hot deck imputation, which is the main source of imputation used in the ACS. Other Census Bureau surveys are already applying these techniques. The 2020 census is using administrative data to reduce non-response follow up efforts and to impute from the same response. The research that led to the decision for the 2020 census motivated us to consider these methods in the ACS. This research that was done matched 2010 census records to administrative data on

Hispanic origin, races, and age.

They found a high match rate among census response and administrative data, with 90 percent or more agreeing. However, the study also compared imputed 2010 census values with administrative data and found some differences. For example, using administrative data resulted in increases in Hispanics and several race categories. Census also appeared to impute more older ages than younger ages when compared to administrative records. These results suggested that using AR in lieu of statistical methods could result in higher data quality. Additionally, since the methodology had already been created and proven for use in the 2020 census, we felt it would be relatively easy to implement into ACS. Therefore, we decided to test integrating administrative records in the ACS edits and imputation procedures.

Our test uses 2016 ACS response data. It incorporates data from previous surveys, the 2010 census, and other administrative sources, which we use to fill in missing response for the demographic characteristics, age, sex, race, Hispanic origin, and place of birth. The research plan is to replace missing responses and then will run the resulting test file, both with the ACS response data and administrative data through our ACS edits and create estimates that we can compare to the published estimates that were published for the 2010 ACS. Unfortunately, high priority projects have delayed our research, and we have not had a chance to run the data through our full edit process.

That research is currently underway, and we've picked it back up, but at this time, the research that I'm presenting today does not account for how the edit procedures may clean up inconsistent administrative data, or how using administrative data could have an impact on how we edit other survey items besides those that are directly included in our test. So this preliminary research discussed today used the 2016 ACS file that had already been

processed through our edits, so the final file. What we did was we replaced allocated values, when there were AR data available, the administrative data. This allowed us to study how the administrative data compared to what we - how we allocated the missing data with using our hot deck approach.

Our administrative data came from four data sources, and we use the following hierarchy for replacement. So initially, we took a look - we looked for data from internal Census Bureau sources that have broad coverage. So first, we turned to the 2010 to 2015 ACS. This data provided more recent data. The ACS samples addresses, not people, and our design makes it impossible for an address to be in sample more than once every five years. So, it is rare for respondents living at an address in our 2016 sample that we used in this test to have a response data from an ACS between 2010 and 2015. However, due to people moving and changing addresses, it is possible to have data, for these respondents. Next, we look for response to 2010 census. If we can't find census data, then we turn to the Social Security Administrative - Administration Numident file. So this file contains transactions recorded for a social security number.

The Numident includes names, date of birth, sex, race, place of birth, parents' names, citizenship status, and date of death information. As a last resort, we check the Census Bureau's Best Race and Hispanic origin file. This file is a composite file that contains a number of federal, state, and third party sources. It employs a method that compares race and ethnicity across a number of these different sources, and when they differ, it chooses one of them based on a best tested strategy. This slide shows the flow of the adaptive design we are simulating. Once we receive ACS returns, we look for response to the test items. If a response is provided, we use that value. If the item was left blank, we look for available administrative data, and we use that value. If we don't have administrative data, then we follow our traditional method and allocate a

value using a hot deck approach. The following research questions were constructed to guide our research.

The first looks for what proportion of ACS respondents did not provide a response for these items. Next, we look at, of those that don't provide a response, how many of these missing values can be filled in with available administrative data? Then, we look at the proportion of the AR values, how they match the ACS allocated values. So we look at match rate between the two. As I mentioned earlier, we had some setbacks, and therefore, we haven't got to research question four or five. We're currently working on these, and putting these test data through our edits.

But I've decided just to include them here, to give you an idea of the other things we hope to learn. So we would like to see whether or not using administrative records changed the overall distribution of the test items when we compare them back to the published estimates, so the impact really of using this data, and then also if there's any effects on the other survey items not included in our test. So we sometimes notice that when we make changes, they have a ripple effect downward, and we want to make sure that this is not the case here. Now, our preliminary, as we get back into this - a preliminary finding suggests that there is minimal impact to the published estimates, or other survey items through this ripple effect. Okay, so now for our preliminary findings. The items included in this test, they have really low missing data rates already.

Most respondents provide answers to the survey items. We see here on this slide that about one percent of the ACS respondents leave the age item blank, less than half a percent of our respondents leave sex missing. It jumps up a little bit to about 1.6 percent of ACS respondents do not provide a response for race, and also, that's the same for Hispanic origin. Now, we look at place

of birth, we find that 6.8 percent of ACS respondents do not provide a response to this item, and for - the reason for the higher response for this is the ACS first we began with creating a roster and asking some of these basic characteristics, such as age, sex, race, and Hispanic origin.

And then we switch over to asking questions about the household structure, and then we go back to asking about all the individual household members. And that is the section where the place of birth is asked, and we do know that sometimes, you know, people break off before getting to this section, or the very beginning of it. But while there is high - these items are highly reported, we still do believe that there's potential for improving - improvements using this data that's already available, and we have an obligation to do so under Title XIII. So, of the missing values, next we look at what percent can be filled in with available administrative data. This slide here shows that the percentages vary by survey item. A smaller proportion of missing response for the age and sex items can be filled in with administrative data, then for the other three items.

So, for age, we have 33.8 percent of the missing values we have available data for, that we could use, for races its about 44 percent. We do better for sex. We have administrative data to fill in over 76 percent of these missing values. Hispanic origin, we have 69 percent. The missing can be filled in, and place of birth is almost 82 percent. Next, we compare the administrative data and - to the ACS allocated values, to see how well they align with one another. Again, these rates vary by survey item. Here, we see that for age, we only - the administrative record value matches the ACS value that we used, that we allocated, only 3.2 percent of the time. When we look - that's looking at individual age ranges.

So when we group the age ranges into some of the groups that we normally

use for published product, we get a better match rate, and about 22 percent of the time, the administrative record value matches how we would have - how we did allocate for ACS. The sex, we do much better. This is not surprising, since there's only two options, male or female, and they match 94 percent of the time. The race is a little over half, at 52.8 percent. That's looking at the individual race groups, which there are hundreds of them. When we combine them into the major race groups, we get a better match rate, - nearly 67 percent.

For Hispanic origin, we do well with about 82.1 percent of the administrative values match the allocated values, and when we look at this - looking at more specific Hispanic origin categories, when we just look at the Hispanic versus non-Hispanic, this jumps up to 94 percent. And for the place of birth items, our value, our allocated value matches the administrative record value at 52.5 percent of the time. Now, we found the misalignment in age very interesting. So, we wanted to dig a little bit deeper. This next slide provides more details into the difference for age. The column in the center here shows that 9.4 percent of administrative record age values are the same as the allocated ACS value, or off by just one year.

Expanding out, we find that 31 percent of administrative record in ACS-allocated ages are within plus or minus five years of one another. This suggests that the remaining approximately 69 percent of administrative records and ACS-allocated ages differ by more than five years. The bar to the left of the center shows the proportion where the ACS allocated an age that's younger than the administrative record age. The bars on the right show the proportion where the ACS allocated an age older than the administrative record age. The distribution is very maybe slightly skewed towards the ACS-allocated age being younger than the administrative record age, but the big takeaway here is the two data sources do appear to be pretty different. Now,

this is - again, I need to point out that this is preliminary research, and it assumes that the administrative record value is consistent with other ACS-reported data, and therefore, it does not change in our editing process. So it will be interesting once we have the final data that is run through, to compare these results again and see what we get.

So, the preliminary conclusions are that the administrative records are available, but there's still a pretty large percent of missing ACS values. We did, however, find there's some low match rates, and this suggests that the ACS edits might not be accurately capturing missing response, and that using administrative record data in lieu of these statistical approaches, may improve the quality of the ACS data. Our next steps are we want to again go back and study these differences in more detail, the AR value and ACS-allocated differences, and we want to finish running our test data through our ACS edits and compare the item distributions with our published distribution. We wanted to determine if using these data to fill in for missing values does have this ripple affect and impacts other items that are not included on the test. Now, the ACS edit and imputation process is still a really huge process, and it takes a great deal of server time and space and effort from ACS staff.

So we're actually currently working with our processing folks to develop a more streamlined approach for running our edits for testing purposes, and therefore, we cannot have to take staff time that are busy with production work, and we can efficiently and quickly test using different administrative data sources in lieu of allocating, and we can expand it to much larger items. And particularly, our larger items are, you know, items that have higher missing data rates.

So, using administrative records to allocate for items with higher missing data rates would really provide a much larger benefit to the ACS programs. There



are a lot of research projects going on related to using administrative data in the ACS, but what we've discussed here is just a few of them. We have a list that keeps growing, of more research projects to do, and the American Community Survey is really dedicated to this work, though we do have to recognize that \$675 billion of federal funds are allocated using the ACS data every year. Therefore, any changes to our process, we really must thoroughly research and approach with caution. We are working hard, and we have lots of work to do.

But our shorter-term goal is to implement the use of the administrative record data for editing, while our longer-term goal would be to, in some instances, completely replace survey items or use them for modeling to substitute. So this slide here shows some tentative target dates for using administrative data in the ACS. Now, these are tentative, and cannot be etched in stone yet. We have things, projects come up. Research sometimes get pushed, but it is targets that we can put out there and try to achieve.

So, since 2016, we have been evaluating administrative sources. We're going to continue these efforts by - and dive deeper into sources that we already have access to now, and also new sources as we acquire them and they become available. We're working hard currently to secure necessary agreements with these other federal agencies, to use these data in our production process, and we have plans to implement the methods that I described today, to use administrative records to impute for some basic demographic items, beginning with our 2021 ACS data release. And then, we will expand to impute for additional items in 2022 releases and beyond.

We are aggressively researching how we can use administrative data for housing items, and we hope that we can use property tax data and remove the acreage item as early as our 2022 data release, and in future years, use

administrative data to supplement survey data for other housing items. In calendar year 2022, we are conducting a field test, which includes changing the last 12 months reference period, and it is used in our income question, so that it is consistent with the calendar year corresponding to IRS tax records. If successful, this change could be implemented. This reference period changes early as the 2025 ACS, and right after that, around that time or at least by 2030, we really hope to use administrative data for income items.

And there's other - lots of research going on in this area. So that's all I have today. And now, I'm going to turn it over to (Leah), to discuss her research.

(Leah Clark): Great. Thanks, (Sandy). So today, I'm going to be talking about the research application that uses linked ACS data and tax records. And before I get started, I want to acknowledge my collaborator, (Jennifer Ortman), who you heard from at the start of the call, and my colleagues in the Center for Economic Studies, (John Voorheis Boar) and (Nick Pharris -Ciurej). So, today I'll be talking about measuring educational attainment by child to family income. Thank you. So, cross-sectional surveys like the ACS offer measures of educational attainment on a national scale. And this means that there's a large sample size, and it's collected annually. But for purposes of today's question, the ACS doesn't contain reliable information about what respondents' family income was when they were in childhood.

Other folks have studied this topic with other data forces that have their own challenges. So for example, longitudinal surveys tend to have all the variables that we're interested in, but smaller sample sizes and fewer cohorts. While using only administrative records -- for example, using only tax records -- researchers are able to calculate near population-level statistics, but they're limited in scope. So maybe you can study college attendance, but you can't assess college completion.

So we're going to link ACS data to IRS forms 1040 data -- so these are, you know, regular tax returns -- to bring the measures of childhood family income into the ACS. And using this link data will produce reliable annual statistics on high school completion, college attendance, and college completion by child's family income. I'm going to be focused on cohorts born between 1983 and 1991. So you can think of this as a large swath of the millennial generation. And because the ACS also contains detailed demographic and geographic characteristics, the force of statistics that can be produced using this much data really represent a large step forward for measuring inequality in educational attainment.

So to define educational outcomes, we're going to be using the ACS question that concern the highest level of educational attainment for our respondents. We'll call - we'll define a high school completed as anyone who finished a regular high school diploma, GED, alternative credential, or some higher degree. A college attended will be anyone who reports having attained some college credit for a higher degree. And today I'll use the definition of college completion that focuses on a four-year degree. So if you have a Bachelor's Degree or higher, I'll consider you a college completer.

So up top here, I've illustrated kind of a life timeline. And so if you imagine someone who's responding to the ACS at age 25, they're going to report what their current household income was - or rather what their income was in the past year. And for people who happen to still live with their parents or for people that they grew up with, that income measure may be an accurate depiction of childhood family income, but for the most part, people are no longer living with their parents at age 25.

So to pinpoint what people's income was when they were in high school -- for

example -- we're going to search for ACS respondents in IRS 1040 data. And so specifically we're going to look to see where they show up as dependents. And form 1040 data and during the years in which they would have turned 15, 16, or 17 years old. I'm going to define childhood family income as the average of those three years of adjusted growth income measures. And then going forward I'll just talk about people in terms of high, middle, and low income. And so we define those categories by ranking respondents by childhood family income within each birth cohort and then splitting that ranking into three equal size groups.

So in total, this links dataset where we're taking ACS respondents from 2006 to 2017 and linking them to IRS form 1040 data, results in the data set of over one million links responded. And this is just restricted to young people who were surveyed at ages 24 to 26. And in this link data, we have childhood family income measures for more than 87% of the population. Next slide, please.

So why am I focused on 24 to 26-year-olds? Educational attainment rises throughout people's 20s, especially when you're looking at people's college completion. So ages 24 to 26 is really the youngest ages that you can look at and expect to see kind of stable rates of college completion reported in the population. Next slide.

So what is the educational attainment of young millennials? So 91% have completed high school. 66% have completed college - or sorry, attend college. And 29% have completed college. But going forward I'm going to talk about conditional rates instead. What I mean by that is when I talk about college attendance I'll talk about college attendance among people who completed high school. So 72% of millennials who completed high school attend college and analogously for college completion, 44% of millennials who attend college complete college. And the reason I'm using this - these conditional

rates is that it will allow us to identify or isolate the inequality that arises at each level of education.

So how does educational attainment vary by family income? So for high school completion, 98% of high-income young people complete high school, but only 84% of low-income young people do. In terms of college attendance, 86% of high-income high school completers attend college, while only 60% of low-income high school completers attend college. And for college completion, the gap is the largest. So 60% of college attenders who are high-income finish college, while only 26% of low-income college attenders actually complete a four-year degree.

Next, I'm going to show similar statistics by race and ethnicity. And so here I'm using a race variable that forces each respondent to have a single race. But in, you know, the next version of these statistics I'll be able to also break out multi-racial respondents. But I'll be - today I'll be talking about respondents who are Hispanic versus those who are non-Hispanic and either white, black, Asian, American Indian, Alaska Native, or Native Hawaiian and Pacific Islander. And a great attribute of this large sample size is that we will be able to talk about outcomes for these - for some of these smaller racial subgroups that are often left out of analyses like these.

So in terms of high school completion - so here I'm plotting high school completion rates for each racial subgroup by childhood family income. And first I want to note that the fact that each of these lines is upward sloping implies that there is an income within each of these subgroups. The one exception to this is for Asian Americans. And so Asian American young people who are low income, 95% of them complete high school. But that's not the case for other racial subgroups. And so in particular, American Indian and Alaska Native young people have much lower rates of high school

completion, and still show large disparities in income.

Turning to conditional college attendance, once again the college attendance of young Asian high school completers stands out much higher than other racial subgroups. But another interesting factor here is that for - among low-income respondents, white, black, and Hispanic high school completers who are low-income report similar rates of college attendance, while American Indian and Alaska Native and Native Hawaiian Pacific Islander young people who finished high school are less likely to attend college.

However, when we turn to college completion, the white/black and white/Hispanic gaps open back up. And here again we see stark income gaps for each racial groups. And that includes Asian Americans, though they do remain more likely to complete college than other racial minorities. And once again, the American Indian, Alaska Native young people have the lowest rates of college completion, conditional on having started college.

So next I'll be able to break these out by year and talk about trends in educational attainment. And while I'm not going to show you regression (inaudible) on the next slide -- just in the interest of time -- I do want to note that I have tested the trends that I'll be talking about and the trends that I'm talking about are statistically significant.

So here's kind of the bright spot to today's presentation, which is that low-income young people have made substantial gains in high school completion over the period of time that we're looking at. And there's evidence that when I look at younger ages -- which I'm not showing you today -- that this growth rate for low-income young people has continued and is in cohort form in the '90s as well. So this growth for low-income young people meant that the income gap in high school completion narrowed by more than one fourth for

cohorts born in the 1980s.

However, turning to college attendance, these gains did not translate. The gains in high school completion did not carry over directly. So low-income high school completers made very modest gains in college attendance during this time period. And because the income gap was so substantial to begin with, the 2% gain for low-income high school completers narrowed the low-income gap by less than one-tenth.

And finally, the income gap during the - for cohorts born in the 1980s contained to widen. So high-income college attenders continued to become more likely to complete college, while for low and middle-income young people who were attending college, completion rates were stagnant.

So just to contextualize these results in the existing literature, we know that the high school graduation rate was stagnant for young people born before 1980. But our results are consistent with (inaudible) study, who has shown rising high school graduation rates in the 1980s, especially for black and Hispanic young people. And we document that that rise was really driven by gains for low-income young people. But something that I want to note here -- though I won't get into in detail -- is that I do find that while low-income black and Hispanic respondents made large gains during this period, low income American Indian and Alaska Native young people did not make gains. And so this data allows us to decompose this growth and show that it's not equally distributed across subgroups.

The income gap in college attendance was rising between the 1960s and 1980s. And so we find kind of stabilization and a slight decline in the 1980s, which is good. But we are seeing continued growth in the income gap in college completion. Which had also grown substantially between the 1960s

and 1980s.

And so in conclusion, linking measures from administrative data to cross-sectional surveys like the ACS opens up many new analysis opportunities. And here the benefits came from being able to create this cross-generational measure of income. I document large and persistent disparities in educational attainment by childhood family income. But we show that there are - have been low-income gains in high school completion. Income gaps vary across racial subgroups, but they're consistently large within racial subgroups. And levels also differ across racial subgroups. So this isn't a question of race or income being the driver. Both are present. And finally, American Indian and Alaska Native young people did not make the same gains. And so with that, I will turn the webinar back over to (Jennifer Ortman). Thanks.

(Jennifer Ortman): Okay. So thank you, (Andy), (Chase), (Sandy), and (Leah). I really enjoyed the opportunity to hear about your research and I hope everyone who joined us by phone enjoyed it as well. They have a couple of final slides to talk about how you can communicate with us. While I walk through those, I'll just remind folks that we have a Q & A that will be happening momentarily. So to get yourselves queued up so the Operator can give you access or the ability to speak, you'll need to dial star one. So I'll just finish up while you all get your phones and start - hopefully, lots of you have lots of questions.

So on this slide, we really with the American Community Survey program. Really enjoy hearing about the stories that folks are telling with our data. So we do encourage you with you're working with ACS data to go to [census.gov/acs/www/share-your-story](https://census.gov/acs/www/share-your-story) -- there's dashes in there -- it's a really fabulous way for us to hear about the myriad of uses of ACS data beyond the things that we know - the federal community is using them for.



As we go on to the next slide, so we have many, many ways for you to continue this conversation. And the next and final slide will have contact information for the presenters, but here is some of the other ACS contact information. Ways to sign up for our works about the ACS program, including information about webinars like this. You can, of course, go to [census.gov/acs](https://census.gov/acs) for all things ACS. We have the census customer service center and the ACS user support e-mail. And of course, we love if you're into social media and posting - if you can use the hashtag #acsdata anytime you're referencing our data to get the word out there.

And so finish up this last slide. And we'll leave this slide up. Certainly, you know, if you don't have time to write down everything you could always just do a quick screenshot on whatever device you're using. But we have a few links to papers that the research that we're doing on administrative data. And also the contact information for our presenters.

So again, you heard from (Andy Keller) about the work to identify and predict vacant unit status, (Chase Sawyer) talked about looking at the differences between the survey responses and administrative data on property value. (Sandy Clark) talked about this very exciting work which hopefully will move into the ACS production process in the not too distant future on editing and imputing demographic characteristic information.

And then last -- of course, not least -- the work (Leah) talked about with linking ACS and IRS data and beyond what we can tell with just that cross-sectional data to looking at linkages of other data sources over the life course to see what we can learn about our population and how cohorts change and what they achieve over time. So that concludes the formal remarks. So I will ask the Operator, is there anybody with a question?

Coordinator: I'm showing no questions at this time, but as a reminder, please press star one and make sure you record your name. One moment to see if we collect questions. And our first question, your line is now open.

(Caller 1): Would you define, please, mailable CAPI?

(Jennifer Ortman): Oh, the acronym CAPI?

(Caller 1): Mailable CAPI.

(Jennifer Ortman): Oh.

(Andy Keller): I can take that, (Jennifer). So those are the cases for which - the (inaudible) that computer-assisted personal instrument (inaudible) so those are the non-response cases (inaudible) that you're able to send a mailing to. They (inaudible) send a mailing to.

(Caller 1): Okay, so does that not include then post office boxes?

(Jennifer Ortman): That would be correct. That would - so post office boxes are not mailable addresses for the ACS. It has to be a street address.

Coordinator: Okay. We're ready for the next question?

(Jennifer Ortman): Yes, please.

Coordinator: Our next question, your line is now open.

(Caller 2): Yes, good afternoon. So my question is as an educator and working in a school district where we have a very transient population, my question is

really based around the tracking of homeless youth. And I wanted to know if there has been any tracking of homeless youth using the demographics from the 2000 to 2010 - the 2010 Census in terms of completion of high school and post-secondary education and whether or not that tracking was done across racial demographics.

(Jennifer Ortman): So I'll give an answer and invite any of the others if anybody has any further information. So the ACS is an address-based sample, so the homeless is not necessarily a population that's going to be represented because we do not have an operation for going out and surveying the homeless population. That's in contrast to the decennial census programs. So the 2000 Census, the 2010, and also the 2020 Census. We do have operations to go out and enumerate the transient populations in the census, but we do not have that in place for the ACS. So that would not be something the ACS would be able to provide data on.

(Caller 2): Okay, thank you.

(Leah Clark): This is (Leah). I'll just note that while we can't use the ACS data to answer that question, you know, there is the potential to link IRS data to census responses, which might let us get some insight into that. And then in the Center for Economic Studies, we're also working on some projects to link education administrative data into our data. And so certainly part of that project one of the things that we're interested in better understanding is the outcomes for kids who we either have trouble identifying impact records or who, you know, the education administrative data designates as homeless. And so while I don't have any research to share with you on that today, it is certainly something that we're thinking about and will be looking at in the future.

(Caller 2): Thank you.

Coordinator: Okay, thank you. Our next question, your line is now open.

(Caller 3): I pass at this time. I have forgotten what I was going to ask. So I'll pass this time. Thank you.

Coordinator: Oh, thank you. Our next question, your line is now open.

(Caller 4): I'll pass, you've already answered the questions about the homeless population. Thank you.

Coordinator: All right, thank you. Our next question, your line is now open.

(Caller 5): Yes. I apologize, I came into the webinar late. Therefore I don't know what all was said, so I'm asking is this recorded to where I can access it again?

(Chase Sawyer): Yes, so - I can answer that. This is (Chase Sawyer). So the information about where you can access the webinar has actually been put in the chat. But if you Google Census Academy, you should be able to find information there about all the Census Bureau's webinars and trainings that are available. And so in about two to three weeks, there should be a transcript, the slides, and a recording of this presentation that you can view.

(Caller 5): Okay. Now, you also mentioned slides several times. And I can't see them. So how do I go about seeing those?

(Chase Sawyer): Yes, so those will actually be part of the recording. We post those recordings on YouTube. The slides were presented via WebEx today. But we'll also provide the - a PDF of the slides as well.

(Caller 5): Okay. All right, thank you.

(Chase Sawyer): Yes, thank you.

Coordinator: Okay, thank you. Our next question, your line is now open.

(Caller 6): Hi, thanks for the very interesting presentation. Two quick questions. One is I see that last Friday the 2020 census folks issued a 26 page paper on the intended uses of administrative records for the 2020 census. So I'm curious how much cost fertilization there's been between the 2020 census effort and the ACS effort. And then the second question has to do with the use of state administrative records. I had to get off the phone for a sec, so you may have answered this. But the (inaudible) Census Bureau asked OMB for permission to renew its ability to get administrative records from state governments, but OMB hasn't said yes yet. And that's terminated as of April 30th, so is that causing a problem for you guys?

(Jennifer Ortman): Thank you for the question. This is (Jennifer). The - to answer the first question, a lot of the research you're seeing here -- in particular, the work on predicting vacant units and also the work to do editing and implications of demographic characteristics -- is work that we are doing in collaboration with folks who are working on the decennial census program. And (Andy Keller) actually is - his primary job duties are working on the 2020 Census. So we very much are working across the programs to identify synergies in terms of research and approach where the programs do have overlap in their operations as well as content for the data collection.

So thank you for bringing that up, because we probably should have highlighted that a bit more. And then I do not have information about access

for the state data or the request to OBM, so I don't know. (Andy) or (Leah), have you heard anything about this?

(Andy Keller): No, I'm not aware of the state data aspect of it.

(Leah Clark): I'm not either...

(Caller 6): (Inaudible).

(Leah Clark): Sorry, you can - I was just going to say that the state admin records that I generally work with to my knowledge are still coming in regularly, but those could be under a different kind of process than - is under a research agreement as opposed to being used for census purposes.

(Caller 6): Thank you.

(Andy Keller): And just to get back to your first question, there's more information about some of the modeling techniques that we've used with administrative data as it pertains to the census. So there (inaudible) general statistics, we have advisory groups that we released the paper to, as well as numerous presentations. So if you want to send me an e-mail, I can point you to some of those presentations or look them up via Google yourself. But there are presentations that are more 2020 census-centric in terms of how administrative records are being used. Not only to identify vacant cases but also for the ideas and imputations and such.

(Caller 6): Okay, terrific. And this is (Andrew Keller)?

(Andy Keller): (Andrew Keller), yes, sir.

(Caller 6): Yes, great. I will e-mail you. Thank you.

(Chase Sawyer): I think with that we have time for one more question. And if you have additional questions you can always reach out to us or [acso.users.support@census.gov](mailto:acso.users.support@census.gov) for more information. Operator, were there any more questions?

Coordinator: Yes, we have four more questions. But I'll take - the final question, your line is now open.

(Caller 7): Yes, how are you today? Asking - can you hear me?

(Jennifer Ortman): Yes, we can hear you.

(Caller 7): Hello? Oh, okay. First of all, excellent presentation. I must confess, I was listening to two webinars at the same time, but yours intrigued me more. I still needed the other information, though. But the question that I had goes back to the homeless question. If we're already counting them during the 5 and 10-year census, why - why can't we use the data because they're most likely - you know, the homeless will be there at least a year afterwards, right? Why couldn't we use that information?

But that really wasn't my real question. My real question was -doggone it. If I want to utilize the homeless, how can I utilize that information to - with what I get from the 5 and the 10- year census? I'm in transportation and so I need to know where they're moving around to to make sure that I'm moving them around. You understand what I'm asking? Hello?

(Jennifer Ortman): Yes, so - this is (Jennifer). I don't have an answer to that question. Does anybody...

(Caller 7): I hope I don't sound stupid. I wasn't trying to sound dumb. But if I can't track my homeless population, how do I track my homeless population I guess is what I'm asking.

(Jennifer Ortman): Sure. Yes, it's a - that's a very important question, I think. And it's - a very important sector of the population. It's not one that the ACS is able to provide information about. But I don't know if any of the other presenters are aware of what kind of data products come out of the decennial census that might be useful for this purpose.

(Caller 7): While you're thinking about that, let me ask one more thing. When you talked about the probability of more high-income people going to college and completing college, how is that economic factor placed in there so that it's not just based on race but it's actually more of an economic question? Because of course if I have money to pay for college, I'm more likely going to pay for it and go ahead and go as opposed to start and use whatever federal money I can get and then will run out of money and then just say, "Okay, I can't afford to go back to school."

(Leah Clark): Hey, this is (Leah). So I think you've absolutely kind of pinpointed why the - why we might expect to see low-income kids who start college not be able to finish it. And so part of the reason I talked about race in the statistics today is just because a lot of data sources that researchers use to look at these questions just don't have enough information or enough data observations to estimate those rates within racial subgroups. And I think what one kind of clear takeaway of these statistics is that income matters kind of regardless of which racial subgroup that you're in.

And so it's not just a question of kind of who lives where and who has access



to which universities. So I'm sure that plays into it. But the income issue is pervasive. And it's especially pervasive in college completion. And so the fact that we've seen great gains in high school completion for low-income kids means that, you know, there has been progress made towards making a high school diploma more accessible and attainable. But the same improvements have not shown up at the post-secondary level yet. And that's obviously a very important - you know, getting a college degree is a huge factor for being able to have higher income in the future and have higher economic mobility.

(Caller 7): Thank you. One more...

(Chase Sawyer): Sorry, this is (Chase). We are kind of over on our time, so I did want to let you know that you can always reach out to us at that e-mail address I mentioned. It's in the chat, but it's [acso.users.support@census.gov](mailto:acso.users.support@census.gov). And so yes, if there are any additional questions, feel free to reach out to us. But I do want to be cognizant of the time. That we are getting - we are over at this point. But yes, I do - I want to thank everyone for - first of all, the presenters today taking the time to present and then also everyone that was willing to join us and hear about this great research being done at the Census Bureau.

Coordinator: And that concludes today's conference. Thank you all for participating. You may disconnect at this time. Speakers, please stand by for post-conference.